Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2021, 8(10):175-180

Review Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Predictive Analytics: An Overview of Evolving Trends and Methodologies

Khirod Chandra Panda [1], Shobhit Agrawal [2]

[1]Asurion Insurance, VA, USA | 0009-0008-4992-3873
[2] Visa, WA, USA| 0009-0000-4957-5575

Abstract This paper provides a concise examination of predictive analytics, a discipline crucial for forecasting future trends by analyzing existing data through statistical and machine learning techniques. Our focus is on the practical applications of predictive analytics across various domains, including finance, healthcare, and risk management, underscoring its role in strategic decision-making. We detail the integral stages of predictive analytics, beginning with the clear articulation of the problem statement, an essential precursor to effective analysis. The gathering and preparation of data follow, emphasizing the need for high-quality, relevant data sets and addressing the intricacies of data cleansing and transformation. The heart of predictive analytics lies in the development and tuning of models tailored to the data and problem at hand. This paper navigates through the selection and application of diverse algorithms, from simple regressions to sophisticated ensemble methods, and highlights the iterative nature of model optimization. Evaluating model performance is pivotal, employing metrics like Mean Squared Error, Precision, Recall, and AUC-ROC to ensure accuracy and reliability. The paper elucidates the significance of these metrics in choosing the most suitable model for deployment. Finally, we discuss the deployment phase, which involves integrating the predictive model into real-world applications, and acknowledge additional steps that may be necessary depending on the problem's context. Concluding with a nod to the future, the paper positions the reader at the starting line, ready to delve into the creation of their predictive models, armed with a solid understanding of the core processes of predictive analytics.

Keywords Analytics, Bigdata, Deep Learning; Machine Learning, Regression

1. Introduction

Predictive Analytics has emerged as the cornerstone of data-driven decision-making in the 21st century. By leveraging historical data, statistical algorithms, and machine learning techniques, predictive analytics enables organizations to forecast future events with an unprecedented level of precision. The roots of predictive analytics can be traced back to the earliest forms of statistical analysis, yet the advent of big data and advanced computational power has exponentially expanded its capabilities and applications [1].

This paper aims to explore the rich tapestry of evolving trends and methodologies that define contemporary predictive analytics. With its multifaceted applications ranging from forecasting market trends to anticipating customer behavior, predictive analytics stands as a revolutionary force across various industries [2]. The historical evolution of this field not only reflects advancements in technology but also shifts in business paradigms, where data has become a strategic asset.

The advent of modern machine learning and the sheer volume of data available for analysis have significantly transformed predictive models. These advancements have enabled finer granularity and greater accuracy in predictions, which have profound implications for strategic planning and competitive advantage.

Furthermore, as predictive analytics becomes more accessible and integrated into business operations, it presents new opportunities and challenges. The ethical considerations, data privacy concerns, and the potential for biases in predictive models are areas of ongoing discourse that merit thoughtful examination.

Predictive analytics has a wide range of application in many domains. Insurance companies collect the data of working professional from a third party and identifies which type of working professional would be interested in which type of insurance plan and they approach them to attract towards its products [3]. Banking companies apply predictive analytics models to identify credit card risks and fraudulent customer and become alert from those type of customers. Organizations involved in financial investments identify the stocks which may give a good return on their investment, and they even predict the future performance of stocks based on the past and current performance. Many other companies are applying predictive models in predicting the sale of their products if they are making such type of investment in manufacturing. Pharmaceutical companies may identify the medicines which have a lower sale in a particular area and become alert on expiry of those medicines [4].This paper provides an overview of the latest trends in predictive analytics, comparing traditional statistical methods with advanced methodologies such as machine learning and deep learning. It also highlights the challenges practitioners face in the current landscape and anticipates future developments in the field. Through an analysis of various case studies and industry applications, the paper will demonstrate the transformative impact predictive analytics has had and will continue to have in the digital era [4]

2. Evolution of Predictive Analytics

Predictive analytics involves several steps through which a system can predict the future based on the current and historical data.

Early Foundations (1950s - 1970s): Predictive analytics traces its origins to the post-World War II period, characterized by pioneering work in statistical modeling and operations research. The development of linear regression models and the introduction of time-series analysis marked the first steps towards data-driven prediction. In the 1960s, the focus shifted to decision theory and Bayesian inference, providing new frameworks for incorporating uncertainty and prior knowledge into predictions

Rise of Computational Power (1980s): The 1980s heralded the age of the personal computer, which dramatically increased computational power and data storage capabilities. The development of the relational database during this era allowed for more structured data collection, leading to more sophisticated analytics. The advent of decision trees and the emergence of the first commercial statistical software packages made predictive analytics more accessible to businesses

Data Explosion and the Internet (1990s): With the proliferation of the internet in the 1990s, data availability surged. This era saw the initial application of neural networks, a class of algorithms inspired by the biological neural networks that constitute animal brains, which could learn from and make decisions based on data. These methods laid the groundwork for what would become machine learning.

Machine Learning and Big Data (2000s): The 2000s were defined by the explosion of big data and significant strides in machine learning, particularly with the development of ensemble methods like random forests and boosting algorithms. These methods improved prediction accuracy by combining the predictions of multiple models. The concept of data mining also came into prominence, focusing on discovering patterns in large datasets

Breakthroughs in Deep Learning (2010s): The 2010s witnessed breakthroughs in deep learning, a subset of machine learning characterized by models that are composed of multiple processing layers. The success of deep learning models, especially in image and speech recognition tasks, was bolstered by the increasing availability of large labeled datasets and improvements in computing power, particularly through the use of Graphics Processing Units (GPUs)

Real-time Analytics and AI Integration (2020s): The present decade is seeing the integration of real-time analytics into business processes. The Internet of Things (IoT) generates a constant stream of data that can be analyzed in real-time, enabling immediate decision-making. Concurrently, advancements in artificial intelligence (AI) are creating systems capable of not just prediction but also prescriptive analytics, which not only forecasts outcomes but also suggests actions to achieve desired results

The Convergence Era: We are currently in an era of convergence, where predictive analytics is not a standalone tool but is integrated with other technologies. Cloud computing facilitates the storage and processing of vast datasets; AI provides the intelligence to make sense of this data, and edge computing enables analytics to be performed at the source of data collection.

The Ethical and Regulatory Horizon: As we advance, the field of predictive analytics is increasingly confronting ethical and regulatory challenges. The formulation of privacy laws like GDPR in Europe, and the discussions around AI ethics, are shaping the development of predictive models to ensure fairness, accountability, and transparency.

3. Current Trade In Predictive analytics

3.1 Machine Learning and AI driven Predictive Analytics

The integration of machine learning (ML) and artificial intelligence (AI) with predictive analytics has been a game-changer. These technologies have reshaped predictive models to be more self-learning and adaptive, capable of handling unstructured data such as images, text, and voice. ML algorithms, particularly supervised learning methods, have become the standard for developing predictive models due to their ability to improve over time with exposure to more data

Deep Learning

A subset of ML, deep learning. [12], has shown remarkable performance in predictive analytics. Its ability to extract features and learn representations from vast amounts of data makes it an ideal choice for complex predictive tasks, such as natural language processing and image recognition. Deep learning's ability to harness large-scale neural networks is proving to be instrumental in solving problems that were previously thought to be intractable.

3.2 Bigdata And Predictive Analytics

The exponential growth of data, also referred to as 'big data,' , has fueled advancements in predictive analytics. The availability of large datasets has led to more accurate and granular predictive models. Moreover, big data technologies have enabled the processing and analysis of these datasets in a distributed and more efficient manner. Frameworks like Apache Hadoop and Apache Spark are used extensively for developing predictive models that can process large volumes of data in parallel.

Realtime analytics

The demand for real-time analytics , has surged, driven by the need for immediate insights and action. Streaming data platforms like Apache Kafka and cloud services from providers such as AWS, Google Cloud, and Azure offer tools that allow businesses to perform predictive analytics in real-time. These tools can analyze data as it's being generated, providing businesses with the ability to make decisions and respond to trends instantaneously.

3.3 Predictive Analytics in Cloud

The cloud has democratized access to predictive analytics by offering scalable, pay-as-you-go services. Cloud-based predictive analytics tools bring advanced analytics capabilities to a wider range of users, eliminating the need for substantial upfront investment in IT infrastructure. Additionally, cloud platforms are increasingly incorporating ML and AI services that simplify the development of predictive models, making these technologies accessible to non-experts.

AI as a Service

Companies like IBM, Amazon, and Google offer AI as a Service, which allows businesses to leverage their cloud-based AI tools for predictive analytics. This trend has enabled businesses to incorporate complex AI algorithms into their processes without the need for in-house AI expertise.

3.4 Democratization Of data

The democratization of data science, driven by the development of user-friendly machine learning platforms, is enabling more people to build and deploy predictive models. Tools such as AutoML and drag-and-drop interfaces have made predictive analytics more accessible, allowing users without deep technical expertise to develop sophisticated models.

Explainable AI

As AI models become more complex, the need for explain ability grows. XAI seeks to make the outputs of AI models more understandable to humans, which is particularly important in fields like finance and healthcare, where understanding the decision-making process is crucial. This trend towards greater transparency helps build trust in AI systems and ensures their decisions can be audited and validated.

3.5 Advanced Prescriptive analytics

Moving beyond prediction, prescriptive analytics uses insights from predictive models to recommend actions. With the integration of optimization algorithms and simulation, predictive analytics is now able to provide actionable recommendations to achieve specific objectives, further enhancing its business value.

Ethical AI and Responsible Analytics:

There's an increasing focus on ethical AI and responsible analytics, ensuring predictive models are free from biases and respect privacy. The development of ethical frameworks and guidelines is becoming a central concern, as predictive analytics becomes intertwined with critical decision-making processes affecting individuals' lives.

4. Predictive analytics Methodology

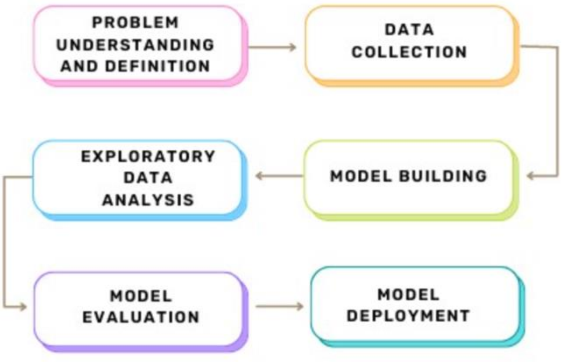
PROBLEM UNDERSTANDING AND DEFINITION
DATA COLLECTION
MODEL BUILDING
EXPLORATORY DATA ANALYSIS
MODEL EVALUATION
MODEL DEPLOYMENT

Figure 1: Methodology

4.1 Problem Understanding and Definition

This marks the beginning of the predictive analysis process, which is crucial because it starts with a clear understanding of the problem to properly shape the solution. When a stakeholder presents a problem, the first step is to gather detailed information about their requirements, available resources, expected deliverables, and the business perspective of the proposed solution.

Sometimes, stakeholders' requirements might not be explicitly clear. In such cases, it is our responsibility to precisely determine what needs to be predicted and whether the prediction addresses the defined problem. The nature of the solution and its results can vary significantly based on how the problem is defined.

Turning a business issue into an analytical question is the most critical step in predictive analysis. It is essential to clearly define what needs to be predicted and what the expected outcome should look like.

4.2 Data Collection

This stage is the most labor-intensive. Occasionally, the necessary data might be provided by the stakeholder, sourced from an external database, or it may require you to extract the data yourself. [6],[7]Often, the initially gathered data might not be adequate for developing the solution, necessitating the collection of additional data from various sources. Consider your level of access to the required datasets.

Given that the success of the predictive model depends entirely on the data utilized, it's crucial to collect the most pertinent data that matches the problem's requirements. When searching for a dataset, keep the following considerations in mind:

- The format of the data
- The time span over which the data was collected.

- The characteristics of the dataset
- Whether the dataset fulfills your specific needs

4.3 Exploratory Data Analysis

Exploratory data analysis (EDA) is used by data scientists [5] to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.

There are four primary types of Data Analysis

Univariate non-graphical. This is simplest form of data analysis, where the data being analyzed consists of just one variable. Since it's a single variable, it doesn't deal with causes or relationships. The main purpose of univariate analysis is to describe the data and find patterns that exist within it.

Univariate graphical. Non-graphical methods don't provide a full picture of the data. Graphical methods are therefore required. Common types of univariate graphics include:

Stem-and-leaf plots, which show all data values and the shape of the distribution.

Histograms, a bar plot in which each bar represents the frequency (count) or proportion (count/total count) of cases for a range of values.

Box plots, which graphically depict the five-number summary of minimum, first quartile, median, third quartile, and maximum.

Multivariate nongraphical: Multivariate data arises from more than one variable. Multivariate non-graphical EDA techniques generally show the relationship between two or more variables of the data through cross-tabulation or statistics.

Multivariate graphical: Multivariate data uses graphics to display relationships between two or more sets of data. The most used graphic is a grouped bar plot or bar chart with each group representing one level of one of the variables and each bar within a group representing the levels of the other variable.

4.4 Model Building

Post EDA, [8] the next logical step is Model building using ML Algorithm Where the target is the dependent variable, and the predicator is the independent variable in the data set

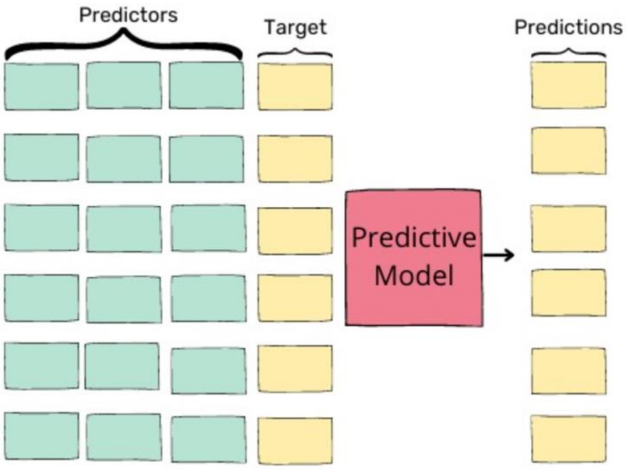

Figure 2: Model Building

Regression algorithms [9] [10] [11] such as Simple Linear Regression, Multi Linear Regression, Decision Tree Regression etc., may be used to get desired results. Such models are used when the target is a numeric feature.

4.5 Model Evaluation

After constructing the model, the subsequent step involves scrutinizing its effectiveness. Assessment across various scenarios and parameters aids in determining the optimal model for addressing the problem at hand. Typically, the model's proficiency is gauged using one or several metrics.

Performance metrics vary based on the type of machine learning model employed.

In the case of regression models, evaluation metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R Squared (R^2 score).

For classification models, the metrics encompass the F2 Score, Confusion Matrix, Precision, Recall, and AUC-ROC.

4.6 Model Deployment

Having constructed, tested, and assessed the model, the phase of presenting it to the stakeholder arrives. Deploying the model entails embedding it into a practical setting for actual use. Deployment can be achieved through incorporation into a software [12] application, melding it with a hardware component, constructing an infrastructure to support the model, or employing the model as a standalone 'data product.'

5. Conclusion

This paper has elucidated the foundational stages of predictive analytics, which are crucial considerations when navigating predictive analytics challenges.

The essential steps include:

- Clearly defining and comprehending the problem at hand
- Gathering and prepping the data set
- Developing suitable models
- Appraising these models to select the most effective one
- Implementing the chosen model in its necessary form

While this summary captures the critical phases, additional procedures may be undertaken contingent on the specificities of the issue.

References

[1]. N. Watson, “History of predictive Analytics: since 1689,” Contemporary Analysis, Apr. 30, 2020. Online Available: https://canworksmart.com/history-of-predictive-analytics/

[2]. R. Dubey, A. Gunasekaran, S. J. Childe, C. Blome, and T. Papadopoulos, “Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture”, British Journal of Management, vol. 30, no. 2, pp. 341–361, Apr. 2019, doi: 10.1111/1467-8551.12355

[3]. Charles Nyce, 2007, “Predictive Analytics White Paper”, American Institute of CPCU/IIA.

[4]. P. Cui et al., "Uncovering and Predicting Human Behaviors", IEEE Intell. Syst., vol. 31, no. 2, pp. 77-88, 2016.